ACADEMIC EXPLANATORY JOURNALISM AND EMERGING COVID-19 SCIENCE: HOW SOCIAL MEDIA ACCOUNTS AMPLIFY *THE CONVERSATION*'S PREPRINT COVERAGE

Alice Fleerackers[1]
Michelle Riedlinger[2]
Axel Bruns[2]
Jean Burgess[2]

[1]Interdisciplinary Studies, Simon Fraser University, Canada
[2]Digital Media Research Centre, Queensland University of Technology, Australia

Journal Citation

Fleerackers, A., Riedlinger, M., Bruns, A., & Burgess, J. (2022). Academic explanatory journalism and emerging COVID-19 science: How social media accounts amplify The Conversation's preprint coverage. *Media International Australia*, 1329878X221145022. https://doi.org/10.1177/1329878X221145022

Academic explanatory journalism and emerging COVID-19 science: How social media accounts amplify *The Conversation*'s preprint coverage

Abstract

This article examines the public communication of COVID-19-related 'preprints' (unreviewed research studies) in a digital media environment. To understand how preprint research flows from preprint server, to media story, to social media audience, we analysed engagement with 'second-order citations'—social media posts linking to media coverage of research—using a sample of 41 media stories published by the research amplifier platform *The Conversation (TC)* that mentioned preprint research during the early months of the pandemic. We applied content analyses to the Facebook and Twitter accounts sharing these stories and analysed the engagement that the posts received. We found that *TC* stories mentioning preprints were shared among a diverse collection of Facebook and Twitter accounts, providing a second layer of social media amplification of preprint research. Still, posts by a small proportion of "elite" actors—people with prominent roles in media and communications, politics, or academia—tended to generate more engagement.

Keywords: The Conversation, COVID-19, research amplifier, social media, journalism, preprints

The onset of the COVID-19 pandemic introduced a "new normal" to almost all aspects of daily life, including science media coverage. In the absence of relevant peer-reviewed research about COVID-19, journalistic coverage of preprints—publicly available research papers that have not yet been peer reviewed—became a key aspect of online discourse about the pandemic (Fraser et al., 2021; Majumder and Mandl, 2020). This wave of heightened media attention to COVID-19 preprints came with benefits and risks. Preprint coverage may have provided timely insights into how to prevent and manage the virus, centering the role of scientists in the process however, several flawed COVID-19-related preprints were covered widely by the media (see Majumder and Mandl, 2020; van Schalkwyk et al., 2020), adding to a growing quantity of misinformation (World Health Organisation, 2020). The risk that preprint media coverage could mislead audiences is recognised by journalists (Authors, 2022); however, journalists are not the only actors shaping public discourse around contingent COVID-19 research. Social media accounts also share, contextualise, and amplify scientific findings, including media stories citing preprint research, by sharing hyperlinks with friends and followers. This study investigates the actors sharing these "second-order citations" (Priem and Costello, 2010) to preprint research on both Facebook and Twitter, using a sample of 41 stories published by the research amplifier platform *The Conversation* during the first months of the pandemic.

The Conversation as a (preprint) research amplifier

A growing body of research is focussing on the emerging roles for academic explanatory journalism initiatives within public communication, mostly notably *The Conversation* (Burgess et al., 2019; Guenther and Joubert, 2021; Hermida and Young, 2019; Marshall, 2015; Osman and Cunningham, 2020). Launched in Australia in 2011 and later expanded to eight additional editions on four continents and one global edition, *The Conversation* (*TC*) has been variously described as a vehicle to promote public intellectualism (Marshall, 2015), a research "amplifier" platform (Osman and Cunningham,

2020), an impact platform (Pooley, 2017), and an inter-media agenda setter (Guenther and Joubert, 2021). The researcher-led stories published in *TC* can be republished—and hence "amplified"—by other media outlets under a Creative Commons licence (Guenther and Joubert, 2021; Osman and Cunningham, 2020), further extending their potential reach. While some *TC* stories focus on disseminating findings of specific research papers, others draw on evidence to provide insights and commentary on issues that are gaining prominence within the news cycle. For example, many of *TC*'s most popular COVID-19 stories focused on debunking prominent conspiracy theories or providing timely, practical "news you can use" to help audiences navigate key challenges of the pandemic (Hermida et al., 2022: 66).

TC's editorial focus on providing evidence-based insight into contemporary issues with direct relevance may explain why many readers report using TC stories to inform both personal and professional decision-making (Zardo et al., 2018). Effectively, its publishing model enables scientists to "take on the role of the journalist and the traditional roles of pitching and writing a story" (Young and Hermida, 2020: 94), while also fulfilling the organisation's mandate of "knowledge mobilization and exchange" (Young and Hermida, 2020: 128). This model was initially developed with peer-reviewed knowledge mobilisation in mind; however, *TC* was one of the most active digital media outlets covering preprints during the early months of the pandemic. TC published 41 stories linking to one or more of the 100 most-cited preprints posted on medRxiv and bioRxiv (two of the most widely used preprint servers) from 1 January to 30 April 2020 (Authors, 2021). Preprints have been key to the academic response to COVID-19 (Aviv-Reuven and Rosenfeld, 2021; Fraser et al., 2021) but can be challenging for journalists to verify and cover effectively (Author et al., 2022). In contrast, researchers writing for *TC* may be considered well-positioned to communicate preprint findings because they possess the content area expertise needed to assess the veracity of the research claims, and because they have access to the editorial support needed to effectively engage non-specialist readers.

Despite oversight by professional journalism editors, studies suggest that *TC's* early pandemic coverage did not consistently follow journalistic guidelines for covering preprints (e.g., Khamsi, 2020; Miller, 2021; Ordway, 2021). Fewer than half of the stories labelled preprint research as a preprint or qualified the findings as preliminary and not peer reviewed, or needing verification (Authors, 2021). We suggested that this may be because *TC*'s audiences comprise mostly university-educated readers who can be presumed to already be familiar with preprints and their limitations and benefits (Authors, 2021). Recently, editions of TC have launched initiatives to engage children and youth *(e.g., Victoria State Government*, 2019). Although TC's audience currently skews young (61% are under 44 years old), about 15% work in academia, 13% in government or policy, 13% in teaching and 10% in healthcare—all of these career areas tend to require at least some higher-level education (*The Conversation*, n.d.).

This assumption of a university-educated audience is challenged by *TC's* amplifier publishing model, as media outlets republishing *TC* content may serve audiences with different levels of education and knowledge of preprints. The possibility of multiple and diverse audiences is further increased by the networked nature of today's digital media environment (Helmond, 2015). In this environment, large social media platforms such as Facebook and Twitter play "an outsized and transformative role in the transmission (distribution) of information" (Box-Steffensmeier et al., 2022: 16), so researchers have had to move beyond a focus on the media stories themselves and into the multiple digital spaces where those stories circulate. Many news consumers no longer actively seek out news but instead encounter it accidentally in their social feeds or streams (Bruns, 2018)—a phenomenon known as "incidental news exposure" (Kligler-Vilenchik et al., 2020: 1026). On social media, citizens and other actors without professional or commercial recognition facilitate such exposure among friends and followers by circulating hyperlinks to media stories—contributing to a "multistep" flow (Ognyanova, 2017) of news information that Hermida (2010: 297) describes as "ambient journalism". Users may be most likely to "discover" news content through the network shares and recommendations of a crowdsourced group of

networked elites—individuals or organisations whose voices become amplified as other accounts in the network reshare, recommend, and engage with their content (Papacharissi and de Fatima Oliveira, 2012).

Second-order citations and the amplification of (uncertain) science

Social media posts that amplify media coverage of research have been termed "second-order citations" (Priem and Costello, 2010), and can be differentiated from "first-order citations", or social media posts that link directly to research papers (or, in this case, preprints). While, in theory, both first- and second-order citations play a similar role in mobilising research knowledge to a wider "public" audience, in practice, the proportion of citizens who engage with first-order citations is often limited (Haustein, 2019; Toupin and Haustein, 2018). Recent research has estimated that as many as 96% of COVID-19 preprints shared on Twitter circulate among a primarily academic audience (Carlson and Harris, 2020). In contrast, people and organisations who share second-order citations on social media (e.g. links to *TC* stories) may be more representative of non-academic research stakeholders, such as activists, businesses, or interested citizens (Lemke et al., 2021). Thus, while a *TC* researcher author may initially disseminate COVID-19-related research through a *TC* story that story may achieve wider, public exposure as social media accounts and crowdsourced elites share links to the story across platforms.

Despite this *potential* for wide public amplification, in practice little is known about the accounts that share *TC*'s preprint stories on social media, or how other accounts in the network engage with (and thus further amplify) these second-order citations. If social media accounts are mainly academic, then second-order citations of preprint research might mirror first-order citations and engage an audience of academic insiders (Carlson and Harris, 2020) with the expertise needed to understand and vet preprint results. In contrast, if other actor groups—such as media professionals, political actors, or citizen

groups—share links to *TC's* preprint coverage, second-order citations to preprints may engage a wider, more diverse public, who may assess preprint research findings as equally credible as peer-reviewed research claims (Wingen et al., 2022) and possibly overestimate the contribution that the reported research can make.

To address these questions, this study adopted a mixed methods approach combining social media analysis and manual content analysis. In this article, we examine the types of accounts (e.g., academic, political, media and communication, citizen and citizen groups) sharing hyperlinks to *TC* stories containing COVID-19 preprint research on Facebook and Twitter. We describe who was sharing these second-order citations and how much engagement these posts and tweets received. By understanding more about the different social media accounts facilitating incidental exposure to preprint research, we shed light on the interconnected roles of amplifier platforms and social media accounts, and the multistep flow of emerging research evidence that can take place in times of crisis.

Method

Identifying The Conversation stories

To identify media stories from *TC* that mentioned contingent COVID-19 science via preprint research, we drew on a dataset from a previous study (Authors). This dataset includes online media mentions of 100 highly circulated COVID-19-related preprints that were posted to the preprint servers bioRxiv and medRxiv during the first four months of the pandemic (see Authors, 2021). The original dataset includes more than 450 online media stories published by 15 online media outlets, from which we selected the 41 stories published by *TC* (henceforth called "*TC* stories", see the Supplementary material, Table 2 for a full list of titles and links). Although we might analyse social media attention to stories by any of these 15 outlets to explore second-order citations of preprint research, we focused on

TC for a number of reasons including: a) our collective interest and funding support to identify new modes of audience engagement that contribute to informed, knowledge-based journalism, b) the uniqueness of TC's research amplifier publishing model (Osman and Cunningham, 2020), which allows TC stories to reach audiences of multiple media outlets at once and thus expands the potential impact of its preprint coverage, c) the expertise of its contributors, who, unlike journalists at other outlets, are ideally positioned to verify preprint findings, and d) the lack of transparency about the unreviewed status of the preprints cited in TC stories (Authors, 2021), which expands the potential for these unvetted findings to be taken out of context and, potentially, mislead users who encounter them incidentally on social media.

Facebook data collection

Using the Facebook public insights tool CrowdTangle, on 26 January 2021 we collected posts that shared the 41 *TC* story links on public Facebook spaces (i.e., public groups, public pages, and public verified profiles). Our sample consisted of Facebook posts mentioning or linking to *TC* stories reporting on preprint research published before 30 May 2020. The final dataset contained 315 public Facebook posts on 266 unique Facebook spaces, sharing 41 stories from *TC* including 55 unique preprint studies (Table 1).

Table 1. Public Facebook posts and tweets sharing links to The Conversation's preprint coverage

Platform	Posts/ tweets	Unique spaces/ accounts	Likes	Other reactions	Shares/ Retweets	Comments/ Replies
Facebook	315	266	15,147	2,274	8,781	3,411
Twitter	1,828	1,325	7,990	n/a	3,390	1,374

Twitter data collection

We collected tweets that included these *TC* story links using Twint, a Python-based Twitter scraping tool that does not rely on Twitter's API (Zacharias and Poldi, 2018). On 6 March 2021, we used Twint to collect tweets and retweets that linked to each of the 41 media stories, along with associated metadata such as the number of likes, retweets, and replies each tweet received. This yielded a total of 1,944 tweets sharing 40 *TC* stories including 55 unique preprint studies (one of the 41 stories had not been shared on Twitter at the time of data collection). As the analysis took place after the initial data collection, we excluded tweets that had since been deleted (n = 7), those that had been shared by accounts that had since been suspended or deleted (n = 92) or changed from being publicly available to private (n = 16) at the time of analysis, and one tweet shared by an account we identified as a bot[i]. This left a final sample of 1,828 tweets from 1,325 unique accounts (some accounts shared more than one tweet linking to a *TC* story; see Table 1 above).

Facebook and Twitter Account Coding

Using content analyses, we categorised the accounts that were amplifying these media stories on both Facebook and Twitter. As Facebook and Twitter have different logics and affordances, the coding approach and coding instruments were tailored to each platform. Following best practices for mass communication research (Lacy et al., 2015), coding for Facebook was performed by one of the authors and a second, independent coder who was not involved in developing the codebook (cf. Strekalova,

2015). The coders used a detailed coding instrument and coding tool[ii] to classify a random subsample of 60 Facebook spaces according to two categories:

1. *Facebook Public Space Type*: whether the space was a public group, a public page, or a public verified profile or unknown (Stocking et al., 2020);
2. *Facebook Public Actor Group*: which actor group the space appeared to be associated with: political actors, business/private corporations, healthcare, civil society organisations, media and communications, academia, or citizens (Gerhards and Schäfer, 2010; Segesten and Bossetta, 2017).

We applied a similar coding approach for Twitter using an adapted version of the codebook[iii]. Again, two coders (one of the authors and the independent coder) independently coded a random subsample of tweets (n = 200) according to the same two categories as Facebook, with minor differences (e.g., instead of categorising accounts as groups, pages, and profiles, they were categorised as groups, individuals, or unknown accounts). For both Facebook and Twitter, coding was performed using the survey tool Cognito Forms[iv] and exported as Excel files for analysis.

Intercoder reliability

We calculated Krippendorff's alpha reliability scores for Facebook and Twitter coding using the standard statistical packages in R that run Krippendorf's alpha with bootstrapping (tidyverse, irr, and kripp.boot). Given that this measure of intercoder reliability is conservative and that the study is exploratory, we set the minimum acceptable level of reliability at .80 (Lacy et al., 2015). We met or exceeded this level of reliability for all codes (Supplementary material Table 3).

After completing intercoder testing, the authors met with the independent coder to discuss discrepancies, identify the sources of disagreements, and refine the codebooks. The rest of the data set (n = 295 Facebook posts, n = 1,628 tweets) was coded by the independent coder. Although this coding was

largely performed by one individual, this coder met regularly with the authors to resolve difficult cases. These discrepancies were discussed until consensus was reached. All descriptive analyses were conducted using Microsoft Excel, R, and Tableau.

Results

Highly shared TC stories containing contingent COVID-19 science

Our analysis revealed that *TC* stories containing contingent COVID-19 research findings received variable attention on Facebook and Twitter, with just a few stories receiving the bulk of engagement and the rest circulating among fewer accounts (see Supplementary material, Figure 1 and Table 2). A story titled "A small trial finds that hydroxychloroquine is not effective for treating coronavirus" (Story #16 in Supplementary material) far outstripped other *TC* stories in terms of the size of the audiences engaging with it on Facebook and Twitter. This high engagement may be, in part, due to the controversy and hope surrounding the drug at the time of data collection. Former US President Donald Trump had been actively promoting the use of hydroxychloroquine as a COVID-19 treatment, despite a lack of evidence supporting the drug's effectiveness (Qiu, 2020). Analyses of tweets mentioning hydroxychloroquine during the early months of the pandemic indicate high levels of polarisation and politicisation within the social media discourse (Marcon and Caulfield, 2021).

Other highly amplified stories were similarly controversial and/or had direct implications for public wellbeing. For example, one of the most engaging Twitter stories (in terms of number of likes, retweets, and replies) was titled "Coronavirus drifts through the air in microscopic droplets—here's the science of infectious aerosols" (Story #40). It described the airborne nature of the virus's spread—a contested topic during the pandemic (Wilson et al., 2020)—and offered tips for reducing chances of infection. Another story that was highly shared on both platforms, titled "5 reasons it's safe for kids to

go back to school" (Story #25), described evidence suggesting that children contracted COVID-19 at lower rates than adults and rarely fell seriously ill. It was published just as Western Australia planned to reopen schools following early closures (Laschon, 2020)—a time when medical professionals were divided about the role of children in transmitting COVID-19 (Murray, 2020). It is also an example of TC's popular "news you can use" style content (Hermida et al., 2022), which may help to explain the high volume of engagement it received.

A story titled, "Evidence obesity is a risk factor for serious illness with coronavirus is mounting, even if you're young" (Story #28) received substantial numbers of Facebook likes and private shares on Facebook, particularly through an infographic combining research findings from six COVID-19-related studies with public health implications. On Twitter, a story titled, "Maybe coronavirus's aggressiveness could be changed by adding or subtracting sugar molecules from its spike protein" (Story #32) was not as widely shared as other *TC* stories but received substantial likes and retweets. Some of this engagement may be attributed to the timeliness and relevance of the story. Published in May 2020—just a few months after the start of the pandemic—the story may have appealed to audiences hungry for potential solutions to mitigating the risks of COVID-19. However, a closer look suggests that relevance and timeliness were not the only factors to consider. In the story, a medical oncologist argued that "among other things, hydroxychloroquine can serve as an oral hypoglycaemic agent, lowering blood sugar", creating a tenuous link between the controversial drug and the severity of the virus. *Fox News* host Laura Ingraham emphasised this link to hydroxychloroquine when she shared this story with her more than 4.2 million Twitter followers on May 7, 2020[v]. Likes, retweets, and comments on this single tweet makes up the bulk of engagement with the story.

Audiences amplifying TC stories containing contingent COVID-19 science

Next, we examined the nature of the accounts that amplified *TC* stories containing contingent COVID-19 science. We found that more Facebook spaces and Twitter accounts outside of the official *TC* accounts amplified these story links in original posts and tweets. Only 43 of the 315 posts shared on Facebook were posted from official Facebook pages belonging to different country editions of *The Conversation*. Collectively, these pages received 10% of all reactions, 30% of comments, and 15% of private shares. On Twitter, official accounts for *TC* accounted for a smaller proportion of the audiences sharing *TC* stories: just 2.93% of the tweets sharing *TC* stories were tweeted by an official *TC* Twitter account. Collectively, these tweets received 2.16% of replies, 8.45% of retweets, and 4.75% of total likes. The relatively high proportion of retweets may be attributed, in part, to the size and nature of *TC'*s followers. While we did not systematically collect follower data for Twitter, we note that several editions of *TC* have sizeable followings, including *TC Australia* (>196,000), *TC UK* (>148,000 followers), *TC US* (>69,000 followers), *TC Africa* (>35,000), and *TC Canada* (>28,000). Only one edition of *TC—TC Global*—had less than 10,000 followers at the time of writing. Beyond size, follower motivations to inform others and amplify news information (Baek et al., 2011; boyd et al., 2010; Holton et al., 2014), may encourage them to retweet links to media stories, such as those first tweeted by *TC.*

Facebook Spaces and Twitter Account Types

Examining these audiences in more detail, we noted differences between Facebook and Twitter in terms of the role that types of accounts played in amplifying these *TC* stories. On Facebook, individual users typically shared *TC* stories on Facebook through groups, accounting for almost half of all spaces in the dataset (n = 148, 46.98%). The next most common public space type were pages (n = 137, 43.49%), followed by verified profiles (n = 18, 5.71%). These proportions remained relatively

stable when considering unique Facebook spaces only (i.e., removing spaces that had shared multiple posts linking to *TC* stories).

About half of the accounts that shared *TC* stories on Twitter belonged to individuals (n = 918, 50.22% of all accounts), rather than groups (n = 228, 12.47%). A large proportion (n = 682, 37.31%) could not be categorised as any account type; these unknown accounts typically included little information in their Twitter bios and used pseudonyms for usernames. They often did not include a profile picture or relied on a generic image. Again, considering unique Twitter accounts, rather than accounts associated with individual posts, did not meaningfully change these proportions.

Actor Groups

To better understand who was sharing the *TC* stories beyond online academic, media, or healthcare communities, we identified the actor groups that Facebook spaces and Twitter accounts affiliated themselves with in their descriptions. We found that citizens and citizen groups (i.e., accounts not associated with any professional actor or organisation category) made up more than half of the Facebook accounts posting *TC* links on Facebook (n = 174; see Figure 1). On Facebook, these citizen accounts were typically posting to groups (rather than pages or verified profiles) and were often geolocated with others (such as in neighbourhood community groups) or engaged with particular issues (such as parent groups or local community groups sharing COVID-19 health information). The next most active Facebook actor groups were Media and Communication (n = 61, 19.37% of all spaces) and Academia (n = 27, 8.57%). Civil society organisations, political actors, healthcare, and private corporations were less common, each comprising about 5% or less of all spaces where *TC* stories were shared. An additional 12 spaces (3.81%) did not provide enough information to be categorised. A similar pattern appeared when considering unique spaces only.

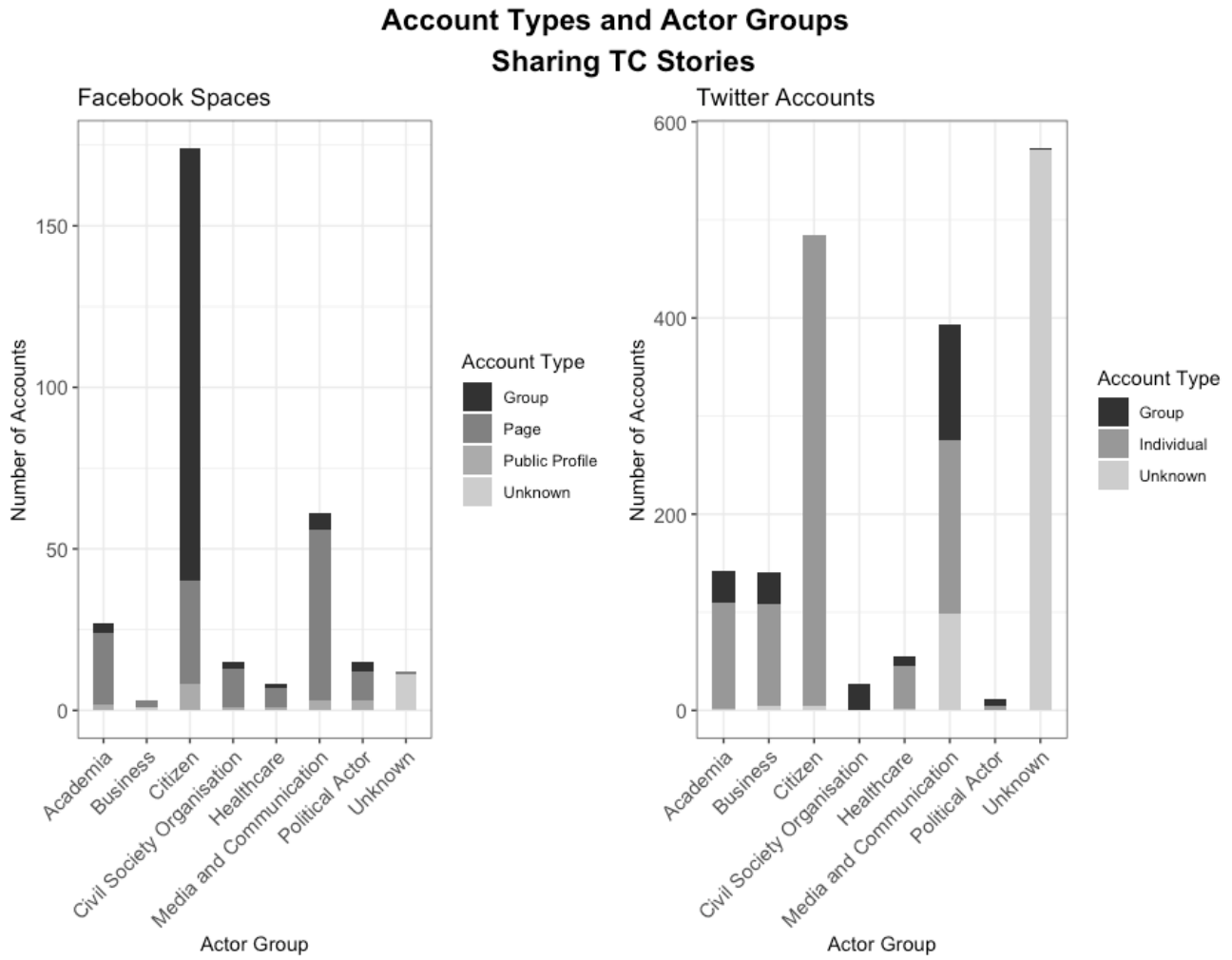

Figure 1. Account types and actor groups sharing TC stories on Facebook and Twitter

It was more difficult to identify the actor groups of Twitter accounts, as many did not provide enough information and were thus coded as "Unknown" (n = 573, 31.35% of accounts, see Figure 1). Citizen was the next most common actor group (n = 485, 26.53%), followed by Media and Communication (n = 394, 21.55%), Academia (n = 142, 7.77%), and Business and Private Corporation (n = 142, 7.71%). Accounts identified as healthcare, political actor, or civil society organisation were not well represented (each ≤ ~3% of dataset). Proportions remained relatively stable when examining only unique accounts.

Engagement with TC stories containing contingent COVID-19 science

To better understand how social media actors were amplifying *TC* stories about COVID-19 preprints, we examined engagement metrics (likes, other reactions, retweets/shares, replies/comments) associated with posts shared by different types of accounts on Facebook and Twitter. While there were similarities across Twitter and Facebook, engagement patterns differed by actor group and platform.

Engagement by account type

On Facebook, posts sharing *TC* stories on pages far outstripped posts shared in other types of spaces in terms of engagement (see Figure 2), although engagement metrics were highly skewed (see Figure 3; and Supplementary material Table 1). Given this skew, we report both mean and median statistics in this section. Medians are commonly used to describe trends when the distribution of data is not symmetrical. Medians are less affected by outliers and skewed data and can sometimes reveal patterns not visible when examining means alone. For example, posts shared to a page received an average of 98.58 likes, 187.64 reactions, 57.15 shares, and 17 comments, but median engagement statistics were far lower (likes = 6, reactions = 4, shares = 3, comments = 0). Posts from three Facebook pages attracted 70% of the total likes and 37% of other reactions, likely because of their high follower counts. These accounts were The Christian Left (political group), KataKita (citizen advocacy group promoting tolerance and anti-Muslim extremism), and Hashem Al-Chali (public science communicator with over 30 million followers on Facebook).

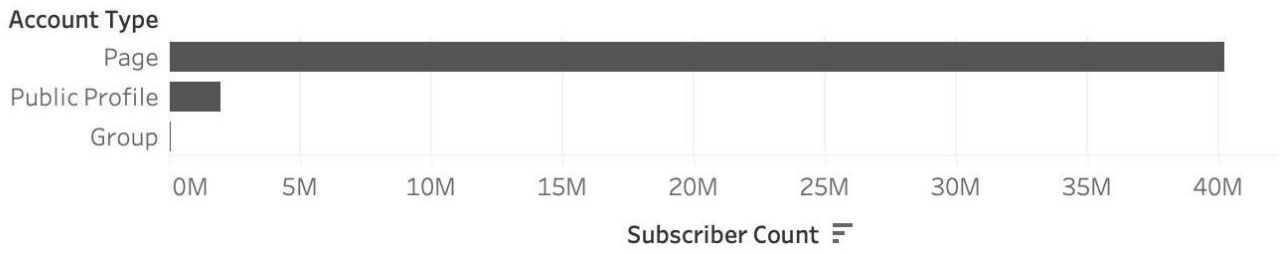

Figure 2. Follower number by Facebook Space Type

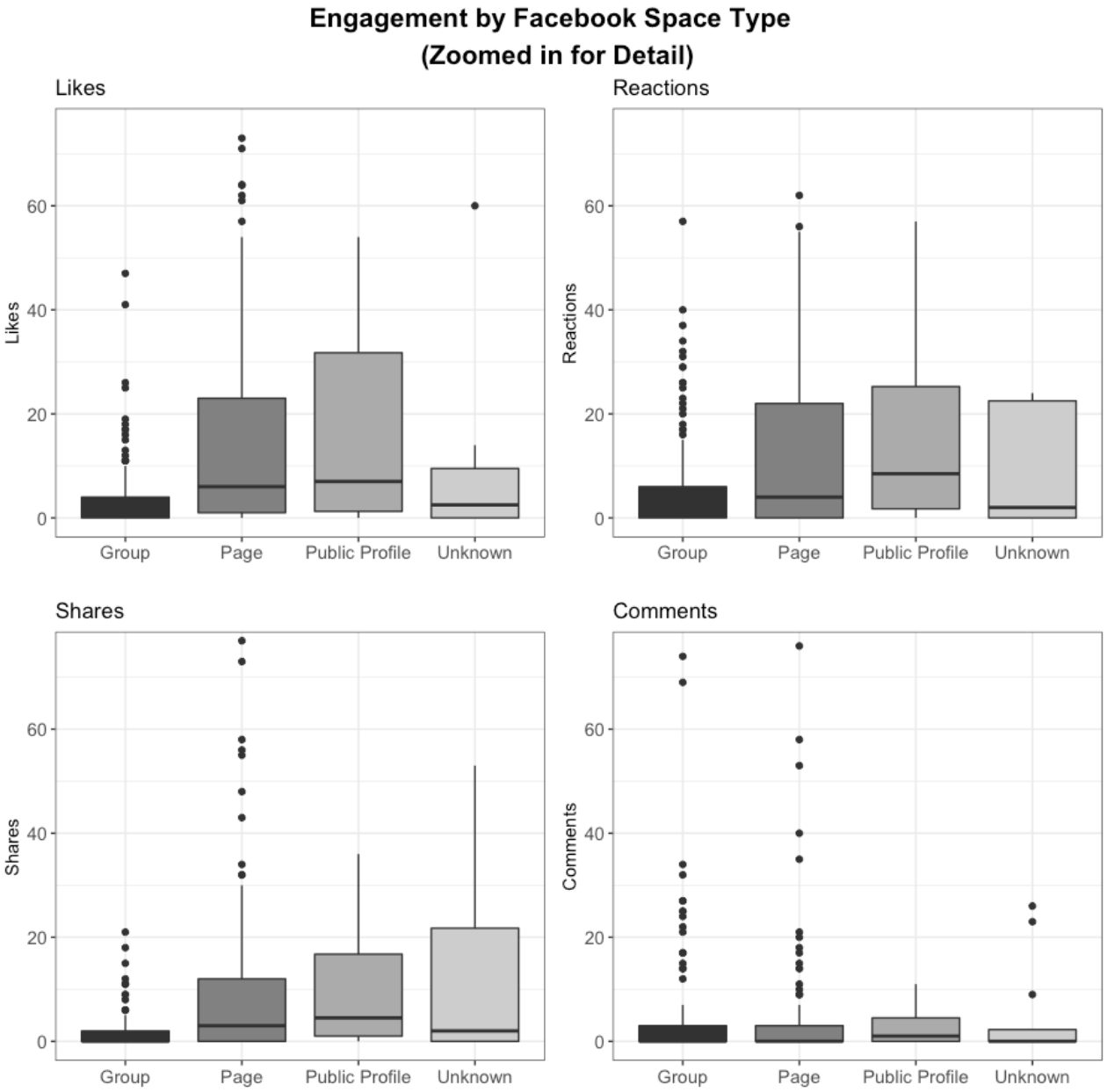

Figure 3. Engagement by Facebook Space Type

Although individual Facebook citizen users shared more links to *TC* stories in groups than in pages and verified profiles, those posts generated far less engagement (in terms of both median and mean across all metrics).[vi]

We also saw different levels of engagement across account types on Twitter, although, again, data were highly skewed, with a few tweets generating extremely high engagement (see Figure 4, and Supplementary material Table 5).

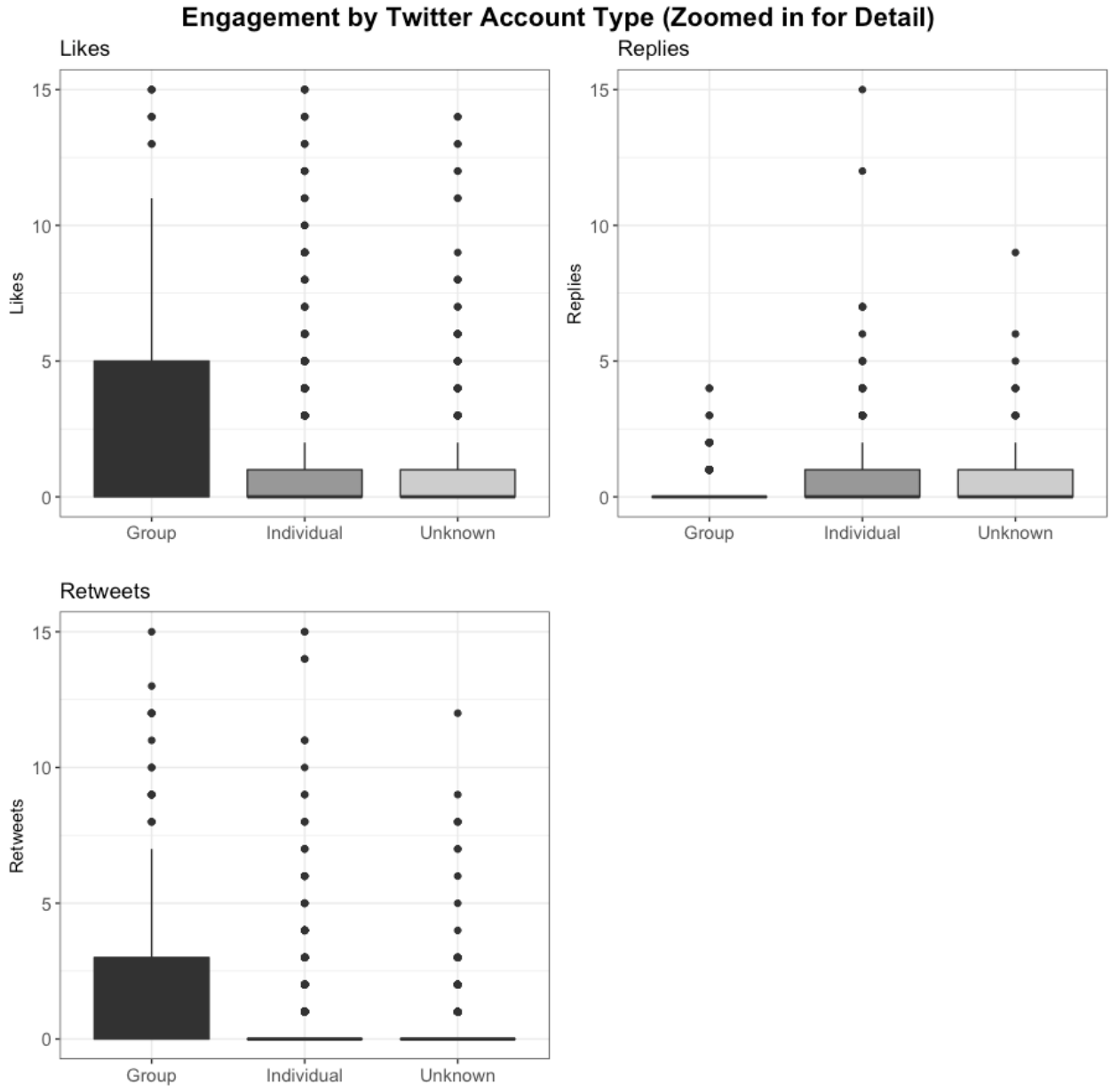

Figure 4. Engagement by Twitter Account Type

When examining average engagement metrics, tweets shared by individual accounts received the most likes, retweets, and shares; however, median counts suggested that tweets by group accounts tended to receive more engagement. Again, it is possible that follower counts played a role in these differences. About a third of the group accounts were associated with Media and Communication organizations (see Figure 1), which tend to have relatively large followings and share content (i.e., links to media stories) that is likely to be retweeted. Thus, *TC* stories shared by these accounts can be expected to receive at least moderate engagement, on average. In contrast, follower counts of individual users may be more variable, with only a few accounts gaining large followings and high visibility. A manual examination of the most and least engaging tweets in our data set provides some support for this assumption. Many high engagement tweets were shared by influential individuals including celebrities (e.g., Kristen Johnston, >323,000 followers); media personalities (e.g., Laura Ingraham, >4,200,000; Rob Crilly, >33,000); science communicators (e.g., Hashem Al-Ghaili, >16,000; Shelly Miller, >19,000; Timothy Caulfield, >79,000 followers); and high-profile business professionals (e.g., Guy Kawasaki, >1,400,000), as well as by ordinary citizens with large Twitter followings. In contrast, tweets that received very little engagement were often shared by individual accounts with fewer than 1,000 followers, with some having less than 15. While a small proportion of individuals sharing *TC* stories could be considered influential crowdsourced elites, the long tail nature of the engagement data suggests that these elites were few.

Again, the high engagement with individual accounts can be at least partially explained by two tweets[vii,viii] authored by influential Fox News host Laura Ingraham, which collectively generated more than 3,000 likes, 1,000 retweets, and 400 comments (see Figure 5). The most engaging tweet included a link to a *TC* story titled “Maybe coronavirus’s aggressiveness could be changed by adding or subtracting sugar molecules from its spike protein,” along with a quote from the story that had been carefully framed to suggest that hydroxychloroquine could help ease the effects of COVID-19.

Figure 5. The two most engaging tweets sharing TC *stories, both by Fox News host Laura Ingraham*

Although tweets from unknown accounts dominated the Twitter data set, these tweets tended to generate fewer likes and retweets, on average, than those shared by individuals and groups. We note that this amplification pattern is characteristic of so-called *spam* accounts—which "may garner many shares because they produce an abundant amount of low-quality posts that each happen to get a little amplification" (Gallagher et al., 2021)—or what has been described as "coordinated inauthentic behaviour" (Keller et al., 2020). Indeed, many of these accounts had abnormal and non-human usernames suggesting inauthentic accounts (Inuwa-Dutse et al., 2018). Tweets from these unknown accounts received slightly more replies than tweets from groups, but we did not systematically examine the nature of these replies. It is possible that the relatively high volume is a sign of other accounts refuting, rather than endorsing, these "low-quality" posts.

Engagement differences among actor groups

Finally, we examined different engagement patterns received by different actor groups on Facebook and Twitter. Like engagement patterns for account types, we found wide variation between actor groups. On Facebook, political spaces contributed 4.44% of the total number of posts; yet these posts received substantially more engagement than those made to spaces of other actor groups (see Figure 6, and Supplementary material Table 6). The average post shared in a political space generated more than 120 likes, 300 reactions, was shared more than 130 times, and received more than 25 comments. In contrast, although Citizen spaces shared the greatest number of posts to *TC* stories, they received far less engagement, with the average Citizen post generating just over 60 likes, more than 100 reactions, over 25 shares, and just under 10 comments.

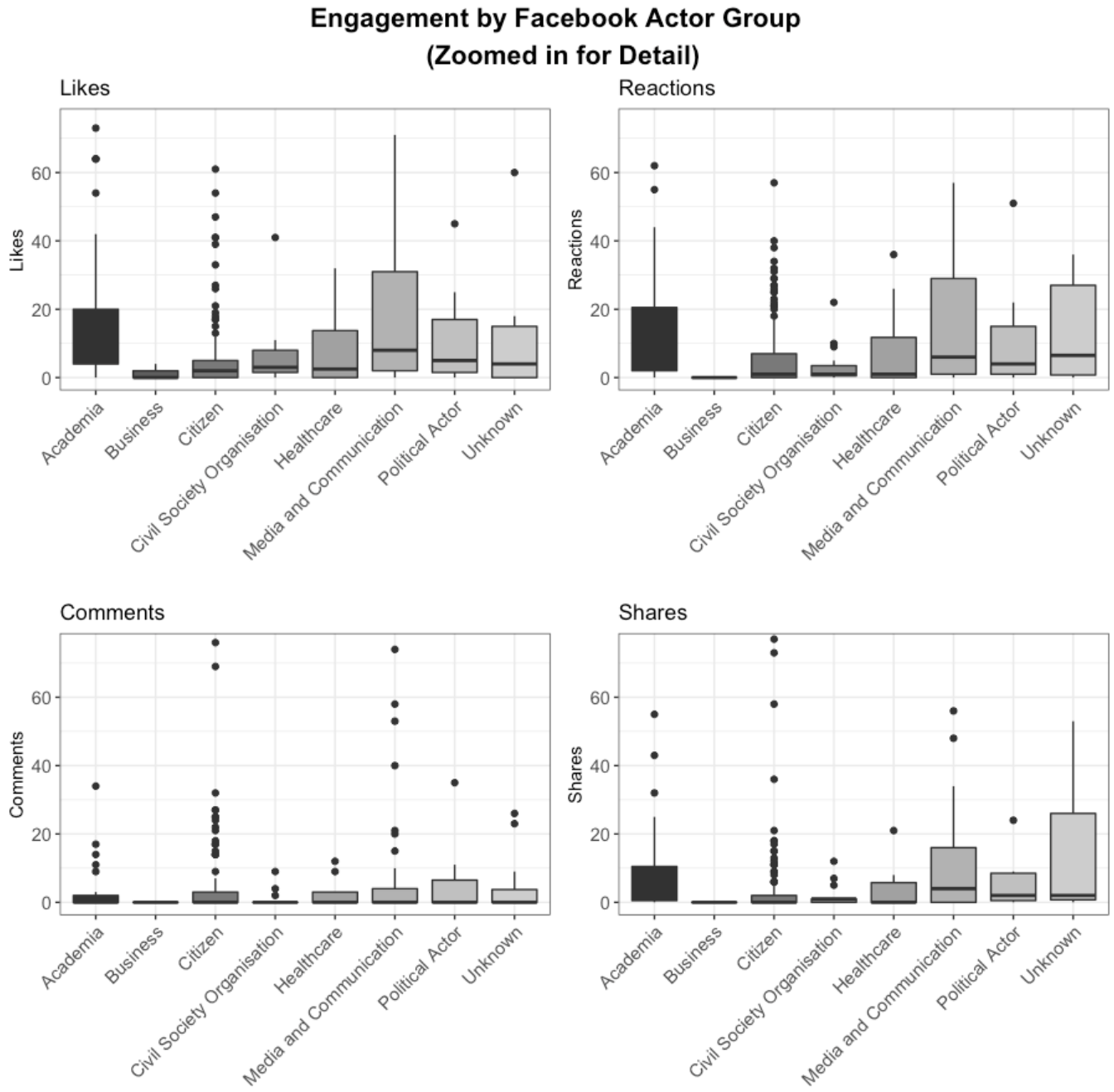

Figure 6. Engagement by Facebook actor group

The Facebook data followed a long-tail pattern of engagement, with a minority of spaces within each actor group receiving the bulk of the engagement, and a far larger group receiving very little. When examining median, rather than average, engagement metrics, patterns of engagement look very different.

For example, while posts shared to political spaces still receive relatively high engagement, they lag behind academic spaces and media and communication spaces in terms of median likes, reactions, and shares. At the median level, no actor group appears to receive many comments.

Twitter actor groups also differed in average engagement their tweets received (see Figure 7, and Supplementary material Table 7), with tweets by some groups generating disproportionately high levels of engagement. Although tweets by Media and Communication accounts made up only 31.35% of the dataset, those tweets generated far more engagement than those by other actor groups. Tweets from academic accounts received similarly high engagement. Again, however, data were highly skewed, with a few highly powerful actors within each group receiving the bulk of the likes, retweets, and replies. While average numbers of likes, retweets, and replies differed wildly across actor groups, when it comes to median levels of engagement, differences between actor groups all but disappeared. Indeed, if median values can be considered more reflective of the norm, then a "typical" tweet from any of the nine actor groups tended to receive almost no engagement.

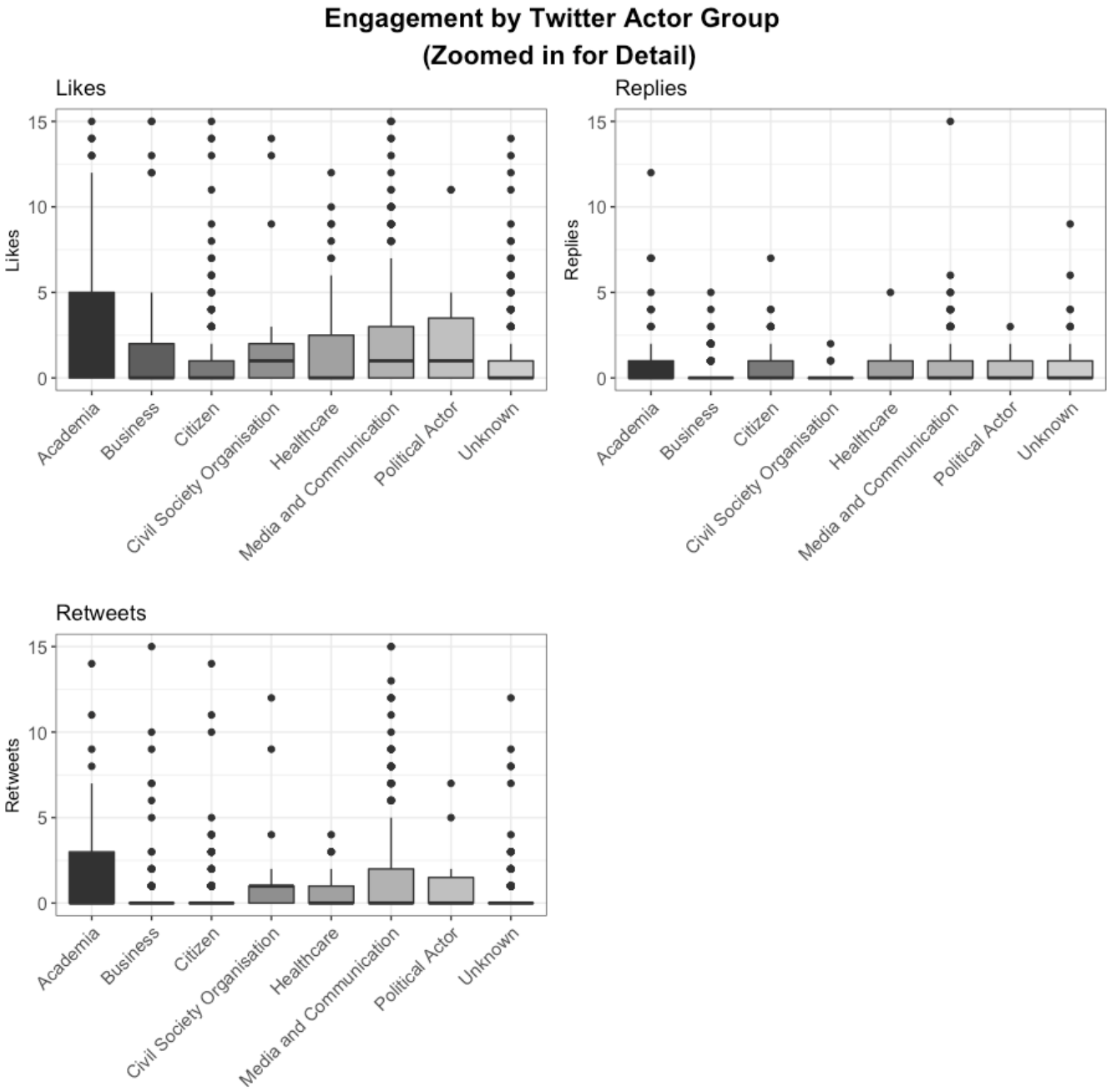

Figure 7. Engagement by Twitter Actor Group

Discussion

This study examined the nature of the social media accounts that amplified media coverage of emerging COVID-19 research through an analysis of "second-order citations" to preprints. We found that *The Conversation (TC*) stories covering COVID-19-related preprints circulated within a diverse

collection of Facebook and Twitter accounts, associated with media organisations, political groups, business professionals, citizens, academics, and more. This diversity of account types amplifying links to *TC*'s coverage of COVID-19 preprints could be seen as both a benefit and a risk. On the one hand, such early access to research may have "brokered" evidence-based knowledge that could have helped digital communities better protect themselves against the deadly virus. On the other hand, platforms such as Twitter and Facebook can encourage the amplification of shallow or misleading content and discourage more nuanced communication about scientific uncertainty (Dunwoody, 2020; Smith et al., 2020; Vosoughi et al., 2018). Indeed, preprints themselves are not inherently problematic for communicating research; however, they may become so if findings are communicated without altering readers to their contingent nature, particularly if those findings eventually fail to hold up to the scrutiny of peer review.

While *TC's* primary audience of university-educated readers may recognise the results of preprint research as potentially subject to change, and understand the processes of scientific review, others engaging with those stories through social media may not (Wingen et al., 2022). Indeed, given that few of the *TC* stories in this study included statements about the unreviewed nature of the preprints they cited (Authors, 2021), most people and organisations sharing links to these stories were likely unaware that they were amplifying preprint research. While many COVID-19 preprints do not appear to change their conclusions dramatically between initial posting and publication in a peer reviewed journal, a sizable proportion undergo major changes (Brierley et al., 2022). To avoid misleading audiences, media outlets brokering knowledge based on preprints need to consider how the results they report could change in future and frame their stories accordingly (Khan, 2021). Such framing may be particularly important when it comes to the headlines, feature images, and excerpts that automatically accompany hyperlinks to media stories when they are shared on social media. Given that many of those hyperlinks

are never clicked (Gabielkov et al., 2016), these components may be the only parts of the story that some social media users ever see.

More broadly, our findings support recent research suggesting that social media accounts that share links to media stories *about* research may better represent society at large than accounts that share links to the research itself (Lemke et al., 2021). The Facebook and Twitter accounts sharing *TC* stories in this study represented a wide range of organisational actor groups; in comparison, social media accounts engaging with links to research articles tend to include a high proportion of academic insiders, such as professors, students, post-doctoral researchers, or publishers (Carlson and Harris, 2020; Toupin and Haustein, 2018). While more research is needed to understand the factors that enabled *TC* stories to reach a broad social media audience, it is possible that *TC*'s innovative publishing model and knowledge mobilisation mission (Hermida and Young, 2019) played a role.

TC has previously been termed an "amplifier" platform (Osman and Cunningham, 2020) because it allows other media outlets to republish stories free of charge. Yet, our findings—alongside those of other scholars (Guenther and Joubert, 2021)—suggest that *TC* can also enable a second layer of amplification to research on social media, particularly in the context of an emerging public health crisis such as COVID-19. Understanding such social media amplification of news in the global "social news media network" (Bruns, 2018) is important, as social media users encounter news incidentally, rather than intentionally, in the form of hyperlinks to media stories shared by friends and followers (Kligler-Vilenchik et al., 2020).

Finally, scholars have argued that social media platforms "profoundly call into question the often-assumed power relationship between science and the public" (Brossard, 2013: 14096) and enable citizens themselves to become "primary drivers of news dissemination and discussion" (Bruns, 2018: 11). While we did find evidence of such grassroots citizen engagement with the science-based media stories in our dataset, our results also suggest that the power structures seen in other forms of journalism

and science communication can persist on social media. Although the accounts that engaged with *TC* stories represented diverse stakeholder groups, these groups were not equally influential in shaping the social media discourse. Social media posts by a small proportion of "elite" actors—such as those working in media and communication, politics, or academia—were shared more often and received more reactions and likes than those shared by ordinary citizens or citizen groups. This finding aligns with research by Gallagher et al. (2021), who similarly identified media and communication actors, political actors, and, to a lesser extent, scientists as powerful crowdsourced elites within Twitter conversations about COVID-19. More research is needed to understand whether social media discourses about issues other than COVID-19 are similarly dominated by a powerful few. As it stands, it is unclear whether the long-tail pattern of engagement found in our study is an artefact of the pandemic context—in which many high-profile actors took to social media to share news about COVID-19—or whether our results reflect wider trends in social media engagement with health and science issues.

Limitations

Findings from this study should be understood alongside several limitations. Our sample of media stories was limited to 41 stories, all of which were published by a single outlet over a period of a few months. This restricted sample allowed us to provide a more detailed view of the social media engagement patterns surrounding these stories at a specific moment in time but does not allow for generalisations about the nature of the social media audiences of all *TC* stories. We were also unable to determine the extent to which this emerging COVID-19 research was shared in private Facebook profiles, pages, groups, or messages, as CrowdTangle does not provide access to this data. In addition, we focused on the social media sharing of *TC* story links themselves; it is possible, given the greater brand reach and recognition of some *TC* story republishers, that the republished *TC* stories would have been shared much more widely on Facebook or Twitter than the original *TC* article itself. Future

research might usefully examine the extent to which the public Facebook spaces and Twitter accounts that shared the 41 media stories in this study differ from accounts where users engage with other second-order citations of research: for example, stories focused on other science-related issues, based on peer-reviewed research rather than preprints, or shared in different languages or sociocultural contexts.

Finally, we did not examine the follower networks of the accounts that shared *TC* stories about COVID-19 preprints; however, it is possible that at least some of these followers included individuals who do not regularly seek out science news. This is particularly likely given that many of the accounts whose posts received the most user engagement—so-called *networked* elites—were also *traditional* elites: journalists, media organisations, and political actors with large public platforms. Future research might usefully build on our work and further examine the degree to which social media amplification of *TC* stories can help facilitate such broad, "incidental" exposure to academic research—particularly beyond the pandemic context.

Conclusion

While social media platforms may offer the potential to destabilise long-standing communication hierarchies, our findings suggest that the reality may not always look quite so democratic. Although citizens can contribute to the wider amplification of emerging science on social media, traditional elites—often with a large number of subscribers or followers—continue to play an important role. Future research is needed to understand how other social media platforms, such as TikTok or Instagram, help shape the multi-step flows of emerging science into the wider, digital public sphere. Such knowledge is sorely needed, as the growth of preprint research and academic explanatory journalism has made it possible for individual researchers to decide when new findings are ready for dissemination—with unknown risks and benefits for the publics who engage with the research.

References

Aviv-Reuven S and Rosenfeld A (2021) Publication patterns' changes due to the COVID-19 pandemic: a longitudinal and short-term scientometric analysis. *Scientometrics* 126(8): 6761–6784. DOI: 10.1007/s11192-021-04059-x.

Baek K, Holton A, Harp D, et al. (2011) The links that bind: Uncovering novel motivations for linking on Facebook. *Computers in Human Behavior* 27(6): 2243–2248. DOI: 10.1016/j.chb.2011.07.003.

Box-Steffensmeier JM, Burgess J, Corbetta M, et al. (2022) The future of human behaviour research. *Nature Human Behaviour* 6(1): 15–24. DOI: 10.1038/s41562-021-01275-6.

boyd danah, Golder S and Lotan G (2010) Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In: *2010 43rd Hawaii International Conference on System Sciences*, Honolulu, HI, January 2010, pp. 1–10. IEEE. DOI: 10.1109/HICSS.2010.412.

Brierley L, Nanni F, Polka JK, et al. (2022) Tracking changes between preprint posting and journal publication during a pandemic. *PLOS BIOLOGY* 20(2): e3001285. DOI: 10.1371/journal.pbio.3001285.

Brossard D (2013) New media landscapes and the science information consumer. *Proceedings of the National Academy of Sciences* 110(Supplement_3): 14096–14101. DOI: 10.1073/pnas.1212744110.

Bruns A (2018) *Gatewatching and News Curation: Journalism, Social Media, and the Public Sphere*. New Edition. New York: Peter Lang Inc., International Academic Publishers.

Burgess J, Bruns A and Osman K (2019) Analysing scholarly contributions to public debate with social media issue mapping: a case study of the Australia Day controversy. *Communication Research and Practice* 5(4): 393–407. DOI: 10.1080/22041451.2019.1689688.

Carlson J and Harris K (2020) Quantifying and contextualizing the impact of bioRxiv preprints through automated social media audience segmentation. *PLOS Biology* 18(9). Public Library of Science: e3000860. DOI: 10.1371/journal.pbio.3000860.

Dunwoody S (2020) Science Journalism and Pandemic Uncertainty. *Media and Communication* 8(2): 471. DOI: 10.17645/mac.v8i2.3224.

Ferrara E (2020) What types of COVID-19 conspiracies are populated by Twitter bots? *First Monday*. DOI: 10.5210/fm.v25i6.10633.

Fraser N, Brierley L, Dey G, et al. (2021) The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS Biology* Dirnagl U (ed.) 19(4): e3000959. DOI: 10.1371/journal.pbio.3000959.

Gabielkov M, Ramachandran A, Chaintreau A, et al. (2016) Social Clicks: What and Who Gets Read on Twitter? In: *Proceedings of the 2016 ACM SIGMETRICS International Conference on*

Measurement and Modeling of Computer Science - SIGMETRICS '16, Antibes Juan-les-Pins, France, 2016, pp. 179–192. ACM Press. DOI: 10.1145/2896377.2901462.

Gallagher RJ, Doroshenko L, Shugars S, et al. (2021) Sustained Online Amplification of COVID-19 Elites in the United States. *Social Media + Society* 7(2): 205630512110249. DOI: 10.1177/20563051211024957.

Gerhards J and Schäfer MS (2010) Is the internet a better public sphere? Comparing old and new media in the USA and Germany. *New Media & Society* 12(1): 143–160. DOI: 10.1177/1461444809341444.

Guenther L and Joubert M (2021) Novel interfaces in science communication: Comparing journalistic and social media uptake of articles published by The Conversation Africa. *Public Understanding of Science*. SAGE Publications Ltd: 09636625211019312. DOI: 10.1177/09636625211019312.

Haustein S (2019) Scholarly Twitter metrics. In: Glaenzel W, Moed H, Schmoch U, et al. (eds) *Handbook of Quantitative Science and Technology Research*. Berlin: Springer. Available at: https://arxiv.org/abs/1806.02201.

Helmond A (2015) The Platformization of the Web: Making Web Data Platform Ready. *Social Media + Society* 1(2): 205630511560308. DOI: 10.1177/2056305115603080.

Hermida A (2010) Twittering the News: The emergence of ambient journalism. *Journalism Practice* 4(3): 297–308. DOI: 10.1080/17512781003640703.

Hermida A and Young ML (2019) From Peripheral to Integral? A Digital-Born Journalism Not for Profit in a Time of Crises. *Media and Communication* 7(4): 92–102. DOI: 10.17645/mac.v7i4.2269.

Hermida A, Varano L and Young ML (2022) The university as a "giant newsroom": Not-for-profit explanatory journalism during COVID-19. In: Ferrucci P and Eldridge II SA (eds) *The Institutions Changing Journalism: Barbarians Inside the Gate*. Routledge.

Holton AE, Baek K, Coddington M, et al. (2014) Seeking and Sharing: Motivations for Linking on Twitter. *Communication Research Reports* 31(1): 33–40. DOI: 10.1080/08824096.2013.843165.

Inuwa-Dutse I, Liptrott M and Korkontzelos I (2018) Detection of spam-posting accounts on Twitter. *Neurocomputing* 315: 496–511. DOI: 10.1016/j.neucom.2018.07.044.

Keller T, Graham T, Angus D, et al. (2020) 'Coordinated inauthentic behaviour' and other online influence operations in social media spaces. *AoIR Selected Papers of Internet Research*. DOI: 10.5210/spir.v2020i0.11132.

Khamsi R (2020) Problems with Preprints: Covering Rough-Draft Manuscripts Responsibly. In: *The Open Notebook*. Available at: https://www.theopennotebook.com/2020/06/01/problems-with-preprints-covering-rough-draft-manuscripts-responsibly/ (accessed 25 June 2021).

Khan S (2021) How The Conversation handles pre-prints. In: *The Conversation*. Available at: http://theconversation.com/how-the-conversation-handles-pre-prints-155063 (accessed 21 April 2022).

Kligler-Vilenchik N, Hermida A, Valenzuela S, et al. (2020) Studying incidental news: Antecedents, dynamics and implications. *Journalism* 21(8): 1025–1030. DOI: 10.1177/1464884920915372.

Lacy S, Watson BR, Riffe D, et al. (2015) Issues and Best Practices in Content Analysis. *Journalism & Mass Communication Quarterly* 92(4): 791–811. DOI: 10.1177/1077699015607338.

Laschon E (2020) 'Soft' opening for WA schools in term two, years 11 and 12 encouraged to go. *ABC News*, 17 April. Available at: https://www.abc.net.au/news/2020-04-17/wa-schools-reopen-amid-coronavirus-students-encouraged-to-attend/12129064 (accessed 20 April 2022).

Lemke S, Bräuer P and Peters I (2021) Does the General Public Share Research on Twitter? A Case Study on the Online Conversation about the Search for a Nuclear Repository in Germany. Glückstadt: Werner Hülsbusch, pp. 94–114. Available at: https://epub.uni-regensburg.de/44940/ (accessed 27 March 2022).

Majumder MS and Mandl KD (2020) Early in the epidemic: Impact of preprints on global discourse about COVID-19 transmissibility. *The Lancet Global Health* 8(5). Elsevier: e627–e630. DOI: 10.1016/S2214-109X(20)30113-3.

Marcon AR and Caulfield T (2021) The Hydroxychloroquine Twitter War: A case study examining polarization in science communication. *First Monday*. DOI: 10.5210/fm.v26i10.11707.

Marshall PD (2015) Understanding the Emerging Contemporary Public Intellectual: Online Academic Persona and *The Conversation*. *Media International Australia* 156(1): 123–132. DOI: 10.1177/1329878X1515600114.

Miller N (2021) News media outlets vary widely in how they cover preprint studies. In: *The Journalist's Resource*. Available at: https://journalistsresource.org/health/how-media-cover-preprint-studies/ (accessed 12 January 2022).

Murray J (2020) WA schools set for soft reopening in term two. *Business News*, 17 April. Perth, Western Australia. Available at: https://www.businessnews.com.au/article/WA-schools-set-for-soft-reopening-in-term-two (accessed 20 April 2022).

Ognyanova K (2017) Multistep flow of communication: Network effects. In: Rössler P, Hoffner CA, and Zoonen L (eds) *The International Encyclopedia of Media Effects*. 1st ed. Wiley, pp. 1–10. DOI: 10.1002/9781118783764.

Ordway D-M (2021) Academic journals, journalists perpetuate misinformation in handling research retractions. In: *The Journalist's Resource*. Available at: https://journalistsresource.org/home/retraction-research-fake-peer-review/ (accessed 27 May 2021).

Osman K and Cunningham S (2020) 'Amplifier' platforms and impact. 62(1): 10.

Papacharissi Z and de Fatima Oliveira M (2012) Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt. *Journal of Communication*: 22.

Pooley J (2017) The Impact Platform. *Parameters*. Humanities Commons. DOI: 10.17613/M65K8K.

Priem J and Costello KL (2010) How and why scholars cite on Twitter. *Proceedings of the American Society for Information Science and Technology* 47(1): 1–4. DOI: 10.1002/meet.14504701201.

Qiu L (2020) Trump's Inaccurate Claims on Hydroxychloroquine. *The New York Times*, 21 May. Available at: https://www.nytimes.com/2020/05/21/us/politics/trump-fact-check-hydroxychloroquine-coronavirus-.html (accessed 20 April 2022).

Segesten AD and Bossetta M (2017) A typology of political participation online: how citizens used Twitter to mobilize during the 2015 British general elections. *Information, Communication & Society* 20(11): 1625–1643. DOI: 10.1080/1369118X.2016.1252413.

Smith M, McAweeney E and Ronzaud L (2020) *The COVID-19 "Infodemic": A Preliminary Analysis of the Online Conversation Surrounding the Coronavirus Pandemic*. 21 April. Graphika. Available at: https://graphika.com/reports/the-covid-19-infodemic/.

Stocking G, Matsa KE and Khuzam M (2020) As COVID-19 Emerged in U.S., Facebook Posts About It Appeared in a Wide Range of Public Pages, Groups. In: *Pew Research Center's Journalism Project*. Available at: https://www.pewresearch.org/journalism/2020/06/24/as-covid-19-emerged-in-u-s-facebook-posts-about-it-appeared-in-a-wide-range-of-public-pages-groups/ (accessed 20 April 2022).

Strekalova YA (2015) Informing Dissemination Research: A Content Analysis of U.S. Newspaper Coverage of Medical Nanotechnology News. *Science Communication* 37(2): 151–172. DOI: 10.1177/1075547014555025.

The Conversation (n.d.) Our audience. Available at: https://theconversation.com/au/audience (accessed 14 October 2022).

Toupin R and Haustein S (2018) A climate of sharing: Who are the users engaging with climate research on Twitter? In: *altmetrics18. Science & The Public: Public Interactions with Science through the Lens of Social Media*, London, UK, 2018.

van Schalkwyk MCI, Hird TR, Maani N, et al. (2020) The perils of preprints. *BMJ*: m3111. DOI: 10.1136/bmj.m3111.

Victoria State Government (2019) The Conversation's Curious Kids comes to the classroom. Available at: https://www.education.vic.gov.au:443/about/news/Pages/stories/2019/stories-curiouskids.aspx (accessed 14 October 2022).

Vosoughi S, Roy D and Aral S (2018) The spread of true and false news online. *Science* 359(6380): 1146–1151.

Wilson N, Corbett S and Tovey E (2020) Airborne transmission of covid-19. *BMJ* 370. British Medical Journal Publishing Group: m3206. DOI: 10.1136/bmj.m3206.

Wingen T, Berkessel JB and Dohle S (2022) Caution, Preprint! Brief Explanations Allow Nonscientists to Differentiate Between Preprints and Peer-Reviewed Journal Articles. *Advances in Methods and Practices in Psychological Science* 5(1). SAGE Publications Inc: 25152459211070560. DOI: 10.1177/25152459211070559.

World Health Organisation (2020) Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. World Health Organisation. Available at: https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation.

Young M-L and Hermida A (2020) *The Conversation Canada: A Case Study of a Not for Profit Journalism in a Time of Commercial Media Decline*. ID 3621685, SSRN Scholarly Paper, 7 June. Rochester, NY: Social Science Research Network. Available at: https://papers.ssrn.com/abstract=3621685 (accessed 23 March 2022).

Zacharias C and Poldi F (2018) TWINT - twitter intelligence tool. Available at: https://github.com/twintproject/twint.

Zardo P, Barnett AG, Suzor N, et al. (2018) Does engagement predict research use? An analysis of The Conversation Annual Survey 2016. *PLOS ONE* 13(2). Public Library of Science: e0192290. DOI: 10.1371/journal.pone.0192290.

[i] For the purpose of this study, we defined "bots" as accounts that explicitly identified as being automated, rather than attempting to categorise accounts that exhibited automated or semi-automated behaviours (Ferrara, 2020)(see Rodríguez-Ruiz et al., 2020; Shevtsov et al., 2020). That is, we coded accounts according to how they presented themselves to other Twitter users rather than attempting to uncover the hidden identities behind them.

[ii] Facebook Codebook to be posted to an online repository post-peer review.

[iii] Twitter Codebook to be posted to an online repository post-peer review.

[iv] https://www.cognitoforms.com/

[v] Follower counts for Twitter accounts were manually identified on July 1, 2022, and may thus over or underestimate the true number of followers associated with those accounts at the time of the original tweets.

[vi] These numbers provide an approximation of the relative influence of these different spaces and accounts; however, CrowdTangle captures the number of followers for accounts at the time of data gathering, which may vary from the number of followers at the time of the original posts; and not all followers will have seen each post from a given space or account.

[vii] https://twitter.com/IngrahamAngle/status/1258414545439264771

[viii] https://twitter.com/IngrahamAngle/status/1254018529680031745